DOCUMENT REFERENCE:
SQ302-002-EN

TEST NODE BRIEFING
Technical information relating to the SamKnows test nodes

August 2013

Contents

1 Important Notice

IMPORTANT NOTICE

Limitation of Liability

The information contained in this document is provided for general information purposes only. While care has been taken in compiling the information herein, SamKnows does not warrant or represent that this information is free from errors or omissions. To the maximum extent permitted by law, SamKnows accepts no responsibility in respect of this document and any loss or damage suffered or incurred by a person for any reason relying on the any of the information provided in this document and for acting, or failing to act, on any information contained on or referred to in this document.

2 SamKnows Test Nodes

In order to gauge an Internet Service Provider's broadband performance at a User's access point, the SamKnows Whiteboxes need to measure the service performance (e.g. upload/download speeds, latency, etc.) from the Whitebox to a specific test node. SamKnows supports a number of "test nodes" for this purpose.

The test nodes run special software designed specifically for measuring the network performance when communicating with the Whiteboxes.

It is critical that these test nodes be deployed near to the customer (and their Whitebox). The further the test node is from the customer, the higher the latency and the greater the possibility that third party networks may need to be traversed, making it difficult to isolate the individual ISP's performance. This is why SamKnows operates so many test nodes all around the world—locality to the customer is critical.

2.1 Test node definition

When referring to "test nodes," we are specifically referring to either the dedicated servers that are under SamKnows' control, or the virtual machines that may be provided to us. In the case of virtual machines provided by Measurement-Lab, Level3, and others, the host operating system is under the control of and maintained by these entities and not by SamKnows.

2.2 Test node selection

The SamKnows Whiteboxes select the nearest node by running round-trip latency checks to all test nodes before measurement begins. Note that when we use the term "nearest" we are referring to the test node nearest to the Whitebox from the point of view of network delay, which may not necessarily always be the one nearest geographically.

Alternatively, it is possible to override test node selection based on latency and implement a static configuration so that the Whitebox will only test against the test node chosen by the Administrator. This is so that the Administrator can choose to test any particular test node that is of interest to the specific project and also to maintain configuration consistency. Similarly, test node selection may be done on a scheduled basis, alternating between servers, to collect test data from multiple test nodes for comparison purposes.

2.3 Test node positioning - on-net versus off-net

It is important that measurements collected by the test architecture support the comparison of ISP performance in an unbiased manner. Measurements taken from using the standardized set of "off-net" measurement test nodes (off-net here refers to a test node located outside a specific ISP's network) ensure that the performance of all ISPs can be measured under the same conditions and would avoid artificially biasing results for any one ISP over another. Test nodes located on a particular ISP's network ("on-net" test nodes), might introduce bias with respect to the ISP's own network performance. Thus data to be used to compare ISP performance are collected using "off-net" test nodes, because they reside outside the ISP network.

However, it is also very useful to have test nodes inside the ISP network ("on-net" test nodes). This allows us to:

- Determine what degradation in performance occurs when traffic leaves the ISP network; and
- Check that the off-net test nodes are performing properly (and vice versa).

By having both on-net and off-net measurement data for each Whitebox, we can have a great deal of confidence in the quality of the data.

2.4 Data that is stored on test nodes

No measurement data collected by SamKnows is stored on test nodes.[1] The test nodes provide a "dumb" endpoint for the Whiteboxes to test against. All measurement performance results are recorded by the Whiteboxes, which are then transmitted from the Whitebox to data collection servers managed by SamKnows.

Note that Measurement-Lab run sidestream measurements for all TCP connections against their test nodes, and publish this data in accordance with their data embargo policy.

1 Note that Measurement-Lab runs sidestream measurements for all TCP connections against their test nodes and publishes these data in accordance with their data embargo policy.

3 Test node hosting and locations

SamKnows test nodes reside in major peering locations around the world. Test nodes are carefully sited to ensure optimal connectivity on a market-by-market basis. SamKnows' test infrastructure utilizes nodes made available by Level3, Measurement-Lab and various network operators, as well as under contract with select hosting providers.

3.1 An introduction to Level3

"Level 3 supports the effort of testing the Internet end user experience, which must include testing the entire Internet ecosystem. This includes testing broadband wire line and wireless access networks, the Internet backbone, middle mile and CDN infrastructures, and the interconnection points between them. Only by testing the entire echo system will we be able to improve the overall quality of Internet services provided to end-users."

Level 3, August 2013

3.2 An introduction to Measurement-Lab

"Measurement-Lab is a research consortium based at the Open Technology Institute (OTI) and dedicated to providing network researchers with the infrastructure they need to conduct open, replicable network measurement.

Measurement-Lab was founded by Vint Cerf, OTI, and a large body of academic researchers, and is supported by over twenty partner organisations, including universities, civil society, and industry organisations. In keeping with a firm commitment to open research and verifiable science, all of the measurement tools running on Measurement-Lab are open source. This means that the methodologies can be reviewed and verified. Similarly, all of the raw measurement data collected against Measurement-Lab test nodes is placed into the public domain and available to anyone wishing to verify a claim or conduct primary research.

Since its launch in 2009, Measurement-Lab has collected over 775 terabytes of network measurement data from across the globe. This number grows daily, all of it free and open to anyone, comprising by far the largest dataset of its kind."

Measurement-Lab, August 2013

3.3 Global test nodes

Level3 has provided SamKnows with 11 test nodes to use for the FCC's Measuring Broadband America Program.[2] These test nodes are virtual servers meeting SamKnows specifications. Similarly, Measurement-Lab has also provided SamKnows with test nodes in various cities and countries for use with the Program's fixed measurement efforts. Measurement-Lab provides location hosting for at least three test nodes per site. Furthermore, SamKnows maintains its own test nodes, which are separate from the test nodes provided by Measurement-Lab and Level3.

Table 1 below shows the locations of the SamKnows test node architecture supporting the Measuring Broadband America Program. All of these listed test nodes reside outside individual ISP networks and therefore are designated as off-net test nodes. Note, that in many locations there are multiple test nodes installed which may be connected to different providers.

Location	SamKnows	Level3	Measurement-Lab
Atlanta, Georgia			✓
Chicago, Illinois		✓	✓
Dallas, Texas		✓	✓
Los Angeles, California	✓	✓	✓
Miami, Florida			✓
Mountain View, California			✓
New York City, New York	✓	✓	✓
San Jose, California		✓	
Seattle, Washington			✓
Washington D.C	✓	✓	
Washington, Virginia			✓

Table 1: Test Node Locations

2 In addition to the test nodes used to support the Measuring Broadband America Program, SamKnows utilizes a diverse fleet of nodes in locations around the globe for other international programs.

SamKnows also has access to many test nodes donated by ISPs around the world. These particular test nodes reside within individual ISP networks and are therefore considered on-net test nodes.

ISPs have the advantage of measuring to both on-net and off-net test nodes, which allows them to segment end-to-end network performance and determine the performance of their own network versus third party networks. For example, an ISP can see what impact third party networks have on their end-users Quality of Experience ('QoE') by placing test nodes within their own network and at major National and International peering locations.

Diagram 1 below shows this set-up.

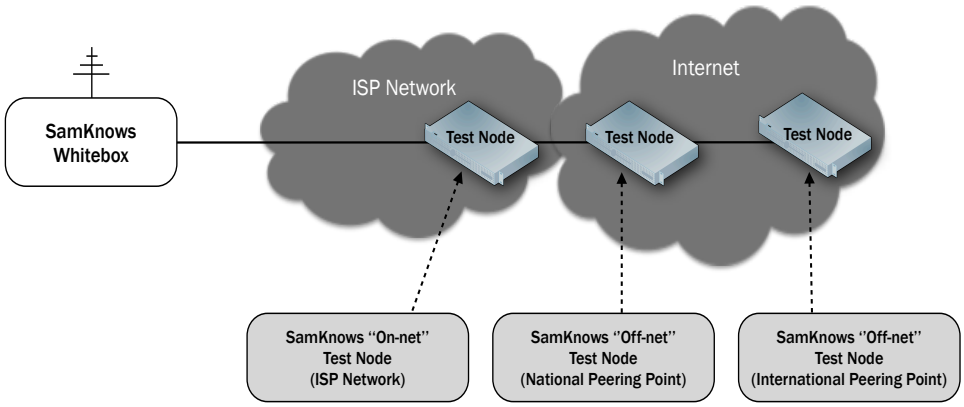

Diagram 1: On-net and Off-net Testing

Both the on-net and off-net test nodes are monitored by SamKnows as part of the global test node fleet. Test node management is explained in more detail within the next section of this document.

4 Test node management

SamKnows test node infrastructure is a critical element of the SamKnows global measurement platform and includes extensive monitoring in place. SamKnows uses a management tool to control and configure the test nodes, while the platform is closely scrutinized using the Nagios monitoring application. System alerts are also in place to ensure the test node infrastructure is always available and operating well within expected threshold bounds.

The SamKnows Operations team continuously checks all test nodes to monitor capacity and overall health. Also included is data analysis to safeguard data accuracy and integrity. This level of oversight not only helps to maintain a healthy, robust platform but also allows us to spot and flag actual network issues and events as they happen. Diagnostic information also supports the Program managers' decision-making process for managing the impact of data accuracy and integrity incidents. This monitoring and administration is fully separate from any monitoring and administration of operating systems and platforms that may be necessary by hosting entities with which SamKnows may be engaged.

4.1 Seamless test node management

SamKnows controls its network of test nodes via a popular open-source management tool called Puppet (https://puppetlabs.com). Puppet allows the SamKnows Operations team to easily manage hundreds of test nodes and ensure that each group of test nodes is configured properly as per each project requirement. Coded in Python, Puppet uses a low-overhead agent installed on each test node that regularly communicates with the controlling SamKnows server to check for updates and ensure the integrity of the configuration.

This method of managing our test nodes allows us to deal with the large number of test nodes without affecting the user's performance in any way. We are also able to quickly and safely make changes to large parts of our test node fleet while ensuring that only the relevant test nodes are updated. This also allows us to keep a record of changes and rapidly troubleshoot any potential problems.

4.2 Proactive test node monitoring

While Puppet handles the configuration and management of the test nodes, Nagios (the most popular online monitoring application) is used by SamKnows to monitor the test nodes. Each test node is configured to send Nagios regular status updates on core metrics such as CPU usage, disk space, free memory, and SamKnows-specific applications. Nagios will also perform active checks of each test nodes where possible, providing us with connectivity information—both via "ping" and connections to any webserver that may be running on the target host.

5 Test node specification and connectivity

SamKnows maintains a standard specification for all test nodes to ensure consistency and accuracy across the fleet.

5.1 SamKnows test node specifications

All dedicated test nodes must meet the following minimum specifications:

- CPU: Dual core Xeon (2GHz+)
- RAM: 4GB
- Disk: 80GB
- Operating System: CentOS/RHEL 6.x
- Connectivity: Gigabit Ethernet connectivity, with gigabit upstream link.

5.2 Level3 test node specifications

All test nodes provided by level3 meet the following minimum specifications:

- CPU: 2.2GHz Dual Core
- RAM: 4GB
- Disk: 10GB
- Operating System: CentOS 6 (64bit)
- Connectivity: 4x1 Gigabit Ethernet (LAG protocol)

5.3 Measurement-Lab test node specifications

All test nodes provided by Measurement-Lab meet the following minimum specifications:

- CPU: 2Ghz 8-core CPU
- RAM: 8GB
- Disk: 2x100GB
- OS: CentOS 6.4
- Connectivity: minimum 1Gbps dedicated upstream

5.4 Test node connectivity

Measurement test nodes must be connected to a Tier-1 or equivalently neutral peering point. Each test node must be able to sustain 1Gbps throughput.

At minimum, one publicly routable IPv4 address must be provisioned per-test node. The test node must not be presented with a NAT'd address. It is highly preferable for any new test nodes to also be provisioned with an IPv6 address at installation time.

It is preferred that the test nodes do not sit behind a firewall. If a firewall is used, then care must be taken to ensure that it can sustain the throughput required above.

5.5 Test node security

Each of the SamKnows test nodes is firewalled using the IPTables linux firewall. We close any ports that are not required, restrict remote administration to SSH only, and ensure access is only granted from a limited number of specified IP addresses. Only ports that require access from the outside world—for example TCP Port 80 on a webserver—would have that port fully open. SamKnows regularly checks its rulesets to ensure that there are no outdated rules and that the access restriction is up to date.

SamKnows accounts on each test node are restricted to the systems administration team by default. When required for further work, an authorized SamKnows employee will have an account added.

6 Test node provisioning

SamKnows also has a policy of accepting test nodes provided by network operators providing that;

- The test node meets the specifications outlined earlier
- Minimum of 1Gbps upstream is provided and downstream connectivity to national peering locations

Please note that donated test nodes may also be subject to additional local requirements.

6.1 Installation and qualification

ISPs are requested to complete an information form for each test node they wish to provision. This will be used by SamKnows to configure the test node on the management system.

SamKnows will then provide an installation script and an associated installation guide. This will require minimal effort from the ISPs involved and will take a very similar form to the package used on existing test nodes.

Once the ISP has completed installation, SamKnows will verify the test node meets performance requirements by running server-to-server tests from known-good servers. These server-to-server measurements will be periodically repeated to verify performance levels.

6.2 Test node access and maintenance

ISPs donating test nodes are free to maintain and monitor the test nodes using their existing toolsets, providing that these do not interfere with the SamKnows measurement applications or system monitoring tools. ISPs must not run resource intensive processes on the test nodes (e.g. packet captures), as this may affect measurements.

ISPs donating test nodes must ensure that these test nodes are only accessed by maintenance staff when absolutely necessary.

SamKnows requests SSH access to the test nodes, with sudo abilities. sudo is a system administration tool that allows elevated privileges in a controlled granular manner. This has greatly helped diagnosis of performance issues with ISP-provided test nodes historically and would enable SamKnows to be far more responsive in investigating issues.

[DOCUMENT ENDS]